working paper
department
of economics

A NOTE ON SPURIOUS BREAK AND REGIME SHIFT IN COINTEGRATING RELATIONSHIP

Jushan Bai

96-13 May 1996

massachusetts
institute of
technology

50 memorial drive
cambridge, mass. 02139

A NOTE ON SPURIOUS BREAK AND REGIME SHIFT IN COINTEGRATING RELATIONSHIP

Jushan Bai

96-13 **May 1996**

96-13

A Note on Spurious Break and Regime Shift in Cointegrating Relationship

by

Jushan Bai[1]

Massachusetts Institute of Technology[2]

May, 1996

[1]Partial financial support is acknowledged from the National Science Foundation under Grant SBR-9414083.

[2]Please send correspondence to Jushan Bai, Department of Economics, E52-274B, MIT, Cambridge, MA 02139. Phone: (617) 253-6217. Email: jbai@mit.edu.

Abstract

The simulation result of Nunes, Kuan, and Newbold suggests that it is possible to estimate a spurious break for a regression model with I(1) disturbances. In this note, we provide a rigorous proof for this phenomenon. We also show that their finding applies to integrated regressors, so that a spurious regression may lead to a spurious break. However, if two integrated processes are cointegrated with a structural change in the cointegrating relationship, the break point can be consistently estimated. The consistency is in terms of the integer index rather than in terms of the sample fraction. This rapid rate of convergence is not attainable for stationary or, more generally, for I(0) regressors. Furthermore, the consistency holds even when magnitudes of breaks are small but do not converge to zero too fast. These consistency results are also obtained for a broken trend model.

Key words and phrases. Spurious break, spurious regression, change point, cointegration, broken trend.

Running head: Spurious Break

1 Introduction

Recently, Nunes, Kuan, and Newbold (1995) pointed out that when the disturbances of a regression model follow an I(1) process there is a tendency to estimate a break point in the middle of the sample, even though a break point does not exist. This phenomenon is called a "spurious break" by the authors and was discovered by a simulation experiment. In this note, we provide a rigorous proof for this phenomenon. Furthermore, we show that a spurious break occurs for I(1) regressors as well, so that a spurious regression may lead to a spurious break.

We then consider the problem in which the dependent variable and the I(1) regressors are cointegrated but the cointegrating relationship undergoes a shift. This is a more general notion of cointegration because the cointegrating vector is not time-invariant. We ask the question whether the break point can be consistently estimated. It is shown that the estimated break point converges quickly to the true break point. The rate of convergence is faster than the corresponding result for I(0) regressors, given the same magnitude of shift.

A structural change in a cointegrating relationship can be a useful model in empirical applications. Cointegration describes a long-run-equilibrium condition. An equilibrium may be disturbed by policy regime changes, resulting in a new equilibrium, so that a different cointegrating vector may be needed to characterize this new equilibrium. A special case is a shift in the mean level of the long-run equilibrium, which can be expressed as a shift in the intercept of a cointegrating-regression model. This shift is exhibited graphically as a change in the "gap" between two cointegrated series.

Although not a concern of this paper, we point out that testing for cointegration which allows for a structural change has been studied by a number of authors; see, e.g., Hansen (1992), Quintos and Phillips (1993), Gregory and Hansen (1996), and Campos, Ericsson, and Hendry (1996). One implication of a structural change in a cointegrating relationship is that one may not be able to reject the null of no cointegration if conventional tests are used, even though a long-run relationship between two series

does, in fact, exist. This situation calls for use of the test statistics proposed by the aforementioned authors.

2 Spurious Break

Consider the model:

$$y_t = \begin{cases} x_t'\beta_1 + \varepsilon_t & t = 1, 2, ..., k \\ x_t'\beta_2 + \varepsilon_t & t = k+1, ..., T \end{cases}$$

Let $\hat{\beta}_1(k)$ be the least squares estimator of β_1 based on the first k observations, and $\hat{\beta}_2(k)$ be the least squares estimator of β_2 based on the last $T-k$ observations, i.e.,

$$\hat{\beta}_1(k) = \Big(\sum_{t=1}^{k} x_t x_t'\Big)^{-1}\Big(\sum_{t=1}^{k} x_t y_t\Big),$$

$$\hat{\beta}_2(k) = \Big(\sum_{t=k+1}^{T} x_t x_t'\Big)^{-1}\Big(\sum_{t=k+1}^{T} x_t y_t\Big).$$

Define the sum of squared residuals for the full sample as

$$S_T(k) = \sum_{t=1}^{k} \Big(y_t - x_t'\hat{\beta}_1(k)\Big)^2 + \sum_{t=k+1}^{T} \Big(y_t - x_t'\hat{\beta}_2(k)\Big)^2$$

and define the break point estimator as

$$\hat{k} = \text{argmin}_{1 \leq k \leq T} S_T(k).$$

Finally, let

$$\hat{\lambda}_T = \min\{\lambda : \lambda = \text{argmin}_{u \in [\underline{\lambda}, \overline{\lambda}]} S_T([Tu])$$

where $0 < \underline{\lambda} < \overline{\lambda} < 1$. The behavior of λ_T is considered for two cases: I(0) and I(1) error processes.

For an I(0) error process, let $Q(\lambda)$ and $R(\lambda)$ be defined as in [A1] and [A3] of Nunes, Kuan, and Newbold (1995, hereafter NKN), respectively. More specifically, $Q(\lambda)$ is the limit of $D_T^{-1/2}(\sum_{t=1}^{[T\lambda]} x_t x_t')D_T^{-1/2}$ for an appropriate scaling matrix D_T, and $R(\lambda)$ is the limit of $D_T^{-1/2} \sum_{t=1}^{[T\lambda]} x_t \varepsilon_t$. The matrix $Q(\lambda)$ is assumed to be positive-definite and strictly increasing. The process $R(\lambda)$ is Gaussian. When x_t is stationary, $Q(\lambda) = \lambda Q$ for some $Q > 0$ and $R(\lambda)$ is a Brownian motion.

In this section, we assume there is no break, i.e. $\beta_1 = \beta_2$. NKN show that (see their Theorem 3.1b)

$$\hat{\lambda}_T \xrightarrow{d} \text{argmax}_{\lambda \in [\underline{\lambda}, \overline{\lambda}]} M(\lambda) \tag{1}$$

where $M(\lambda)$ is a stochastic process given by

$$M(\lambda) = R(\lambda)' Q(\lambda)^{-1} R(\lambda) + [R(1) - R(\lambda)]'[Q(1) - Q(\lambda)]^{-1}[R(1) - R(\lambda)]. \tag{2}$$

For an I(1) error process ε_t, we assume that $T^{-2} \sum_{t=1}^{[T\lambda]} \varepsilon_t^2 \Rightarrow c \int_0^\lambda W^2(u) du$ with $W(u)$ being a standard Wiener process, and $c > 0$ a constant. Further assume (see [A3′] in NKN[1]), for some $\alpha > 0$,

$$T^{-\alpha/2} D_T^{-1/2} \sum_{t=1}^{[T\lambda]} x_t \varepsilon_t \Rightarrow G(\lambda) \tag{3}$$

where $G(\lambda)$ is Gaussian process. NKN prove that

$$\hat{\lambda}_T \xrightarrow{d} \text{argmax}_{\lambda \in [\underline{\lambda}, \overline{\lambda}]} M^*(\lambda) \tag{4}$$

where

$$M^*(\lambda) = G(\lambda)' Q(\lambda)^{-1} G(\lambda) + [G(1) - G(\lambda)]'[Q(1) - Q(\lambda)]^{-1}[G(1) - G(\lambda)]. \tag{5}$$

Examining (2) and (5) we find that, whether the error process is I(0) or I(1), the results are essentially the same. Namely, in the absence of a break, the estimated break point $\hat{\lambda}_T$ is a random variable with support in $[\underline{\lambda}, \overline{\lambda}]$. Not much further can be said on the compact interval $[\underline{\lambda}, \overline{\lambda}]$. However, for $I(0)$ error process ε_t, NKN further prove that $M(\lambda) \to \infty$ as $\lambda \to 0$ or 1, thus $\hat{\lambda}_T \to \{0, 1\}$, if $\underline{\lambda} \to 0$ and $\overline{\lambda} \to 1$; also see Andrews (1993). In their Remark 1 (p. 742), NKN pointed out that they were unable to characterize the limiting behavior of $M^*(\lambda)$ for λ near 0 or 1. Through simulation, they find that $M^*(\lambda)$ behaves differently from $M(\lambda)$. More specifically, $M^*(\lambda)$ does not diverge to infinity as λ decreases to zero or increases to 1.

[1] Their original assumption is stated in terms of y_t rather than ε_t, which applies to y_t being I(1). The current form allows y_t to depend on deterministic regressors as well as on an additive I(1) error process.

In the following, we shall prove that $M^*(\lambda)$ is a well defined process on $[0,1]$ and is uniformly bounded in probability over $[0,1]$. Note that $M^*(\lambda)$ is the limiting process of $T^{-\alpha}M_T^*([T\lambda])$, where α is defined in (3) and

$$M_T^*(k) = \Big(\sum_{t=1}^{k}\varepsilon_t x_t\Big)'\Big(\sum_{t=1}^{k}x_t x_t'\Big)^{-1}\Big(\sum_{t=1}^{k}x_t\varepsilon_t\Big) + \Big(\sum_{t=k+1}^{T}\varepsilon_t x_t\Big)'\Big(\sum_{t=k+1}^{T}x_t x_t'\Big)^{-1}\Big(\sum_{t=k+1}^{T}x_t\varepsilon_t\Big) \tag{6}$$

We shall assume that $\alpha \geq 2$ because this is true when x_t contains a nonzero mean regressor (e.g., a constant, or a trend). When x_t is a $I(0)$ process with zero mean, it is possible that $\alpha = 1$. This case is not considered in this paper.

Theorem 1 *For the α defined in (3), assume $\alpha \geq 2$. We have*

$$\sup_{\lambda\in(0,1)} M^*(\lambda) = O_p(1). \tag{7}$$

Proof of Theorem 1. For an arbitrary vector z and an arbitrary projection matrix P, we have $z'Pz \leq z'z$. Apply this inequality to $M_T^*(k)$ to obtain

$$M_T^*(k) \leq \sum_{t=1}^{T}\varepsilon_t^2 \qquad \text{for all } k \in [1,T]. \tag{8}$$

Since $\alpha \geq 2$, we have $T^{-\alpha}M_T^*(k) \leq T^{-2}\sum_{t=1}^{T}\varepsilon_t^2$ for all $k \in [1,T]$. Moreover, because $T^{-2}\sum_{t=1}^{T}\varepsilon_t^2$ has a limit, $T^{-\alpha}M_T^*(k)$ is uniformly bounded in probability. Thus its limit, $M^*(\lambda)$, is uniformly bounded in probability for $\lambda \in (0,1)$. $\square$

To rule out the possibility that $\hat{\lambda}_T \to \{0,1\}$, we need to further examine the behavior of $M^*(\lambda)$ for λ near 0 and 1. Strictly speaking, $M^*(\lambda)$ is not defined yet at $\lambda = 0$ and $\lambda = 1$. As the limit of $M^*(\lambda)$ when $\lambda \to 0$, $M^*(0)$ should be defined as

$$M^*(0) = G(1)'Q(1)^{-1}G(1) \tag{9}$$

which is obtained from (5) by taking $G(\lambda) = 0, Q(\lambda) = 0$, and $G(\lambda)'Q(\lambda)^{-1}G(\lambda) = 0$ for $\lambda = 0$. Note that the term $G(\lambda)'Q(\lambda)^{-1}G(\lambda)$ is the limit of the first term of (6) on the right hand side divided by $T^{-\alpha}$. Now

$$T^{-\alpha}\Big(\sum_{t=1}^{k}x_t\varepsilon_t\Big)'\Big(\sum_{t=1}^{k}x_t x_t\Big)^{-1}\Big(\sum_{t=1}^{k}x_t\varepsilon_t\Big) \leq T^{-\alpha}\sum_{t=1}^{k}\varepsilon_t^2 \leq T^{-2}\sum_{t=1}^{k}\varepsilon_t^2$$

which converges to zero in probability for any given k, or for $k = [T\lambda]$ with $\lambda \to 0$. It follows that $G(\lambda)'Q(\lambda)^{-1}G(\lambda) \to 0$ in probability as $\lambda \to 0$. Thus the definition of (9) is the limit of $M^*(\lambda)$ as $\lambda \to 0$. Similarly, we can define, as the limit of $M^*(\lambda)$ when $\lambda \to 1$, $M^*(1) = M^*(0)$. We next show that the maximum of $M^*(\lambda)$ is not attained at 0 or 1.

Theorem 2 *(i) With probability 1,*

$$M^*(0) = M^*(1) \leq M^*(\lambda), \qquad \textit{for every } 0 < \lambda < 1. \tag{10}$$

(ii) If $G(\lambda)$ has a nonsingular covariance function, then with probability 1

$$M^*(0) = M^*(1) < M^*(\lambda), \qquad \textit{for every } 0 < \lambda < 1. \tag{11}$$

To prove Theorem 2, we need the following lemma.

Lemma 1 *For arbitrary positive-definite matrices A and B with $A > B$ $(p \times p)$, and arbitrary vectors x and y $(p \times 1)$, we have*

$$x'A^{-1}x - y'B^{-1}y - (x-y)'(A-B)^{-1}(x-y) \leq 0. \tag{12}$$

Proof of Lemma 1: Define the matrix

$$H = \begin{pmatrix} (A-B)^{-1} - A^{-1} & -(A-B)^{-1} \\ -(A-B)^{-1} & (A-B)^{-1} + B^{-1} \end{pmatrix}. \tag{13}$$

It suffices to prove H to be positive-semidefinite because the left hand side of (12) is equal to $-z'Hz$ for $z' = (x', y')$. Let $D = (A-B)^{-1} + B^{-1} > 0$. Let C be a matrix with the first p rows $(I, (A-B)^{-1}D^{-1})$ and second p rows $(0, I)$. Using the identity

$$(A-B)^{-1} - A^{-1} = (A-B)^{-1}D^{-1}(A-B)^{-1}$$

to obtain

$$C'HC = \text{diag}(0, D) \geq 0.$$

Thus $C'HC$ is positive-semidefinite, so is H because C has full rank. This proves the lemma. $\square$

Proof of Theorem 2. The inequality $M^*(0) \leq M^*(\lambda)$ is equivalent to

$$G(1)'Q(1)^{-1}G(1)-G(\lambda)'Q(\lambda)^{-1}G(\lambda)-[G(1)-G(\lambda)]'[Q(1)-Q(\lambda)]^{-1}[G(1)-G(\lambda)] \leq 0.$$

Clearly, part (i) of Theorem 2 follows from Lemma 1 by letting $A = Q(1)$, $B = Q(\lambda)$, $x = G(1)$, and $y = G(\lambda)$. Next, consider (ii). Let $A = Q(1)$ and $B = Q(\lambda)$ and let H be defined in (13). Then $M^*(0) < M^*(\lambda)$ is equivalent to $-\xi' H\xi < 0$, where $\xi = (G(1)', G(\lambda)')'$. Let Γ be an orthogonal matrix such that $\Gamma' H\Gamma = \text{diag}(\lambda_1, ..., \lambda_{2p})$ with $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_{2p}$, where $\lambda_i' s$ are the eigenvalues of H. Since $H \geq 0$ and $H \neq 0$, the maximum eigenvalue of H is positive. It follows that

$$-\xi' H\xi = -(\Gamma\xi)'\text{diag}(\lambda_1, ..., \lambda_{2p})\Gamma\xi \leq -\eta^2\lambda_1$$

where η is the first component of $\Gamma\xi$. When $G(\lambda)$ has a nonsingular covariance matrix, so does ξ. Thus $\Gamma\xi$ is a vector of normal variables with a nonsingular covariance matrix, implying $-\eta^2\lambda_1 < 0$ with probability 1 because $P(\eta^2 = 0) = 0$. That is, $-\xi' H\xi < 0$ with probability 1. □

The above analysis applies to I(1) regressors as well, which is not considered by NKN. Let

$$y_t = x_t'\beta + \varepsilon_t$$

where both x_t and ε_t are I(1), so that we have a spurious regression. Assume

$$T^{-2}\sum_{t=1}^{[T\lambda]} x_t x_t' \Rightarrow Q(\lambda) \tag{14}$$

where $Q(\lambda)$ is a stochastic matrix with $Q(\lambda) > 0$ (a.s.) and $Q(\lambda)$ is strictly increasing, i.e. $Q(u) - Q(v) > 0$ with probability 1 for $v < u$. Also assume

$$T^{-2}\sum_{t=1}^{[T\lambda]} x_t \varepsilon_t \Rightarrow G(\lambda) \tag{15}$$

where $G(\cdot)$ is a vector of random processes, and for each λ, $G(\lambda)$ possesses a density function and a nonsingular covariance matrix.

Theorem 3 *The results of Theorem 1 and Theorem 2 apply to I(1) regressors satisfying (14) and (15), so that spurious regression leads to a spurious break.*

Proof of Theorem 3. First, (14) and (15) imply (3) with $D_T = T^2 I$ and $\alpha = 2$. The proof of (7) under the new setting is identical to the previous proof because inequality (8) is a pure mathematical inequality and holds for arbitrary x_t. To prove (10), we only need to note that if A, B, and $A - B$ are stochastic matrices that are positive-definite with probability 1, then inequality (12) holds with probability 1. The rest of the proof is virtually identical to that of Theorem 2. $\square$.

It remains an open question whether a spurious break arises when y_t and x_t are cointegrated. We conjecture that a spurious break will not occur because an I(1) error process is responsible for its occurrence.

3 Regime Shift in Cointegrating Relationships

In the previous section we show that two integrated processes that are not cointegrated may give rise to a spurious break. What happens when the two processes are cointegrated but the cointegrating relationship undergoes a shift? Can the shift point be consistently estimated in the presence of I(1) regressors?

The issue of a structural change in cointegrating relationships is of considerable interest. Cointegration describes a system's long-run equilibrium condition. A system may have multiple long-run equilibria with an occasional shift from one equilibrium to another. A structural change model allows us to describe such a system. Consider,

$$y_t = \begin{cases} \alpha_1 + \gamma_1 x_t + \varepsilon_t & t = 1, 2, ..., k_0 \\ \alpha_2 + \gamma_2 x_t + \varepsilon_t & t = k_0 + 1, ..., T \end{cases} \tag{16}$$

where $x_t = x_{t-1} + e_t$ with $x_0 = 0$ and $\mathrm{Var}(e_t) > 0$. We assume ε_t are e_t are I(0) linear processes such that $\varepsilon_t = \sum_{j=0}^{\infty} a_j \eta_{t-j}$ and $e_t = \sum_{j=0}^{\infty} b_j \xi_{t-j}$, where η_t and ξ_t are i.i.d. sequences with finite $4 + \delta$ ($\delta > 0$) moments, and $\sum_j j|a_j| < \infty$ and $\sum_j j|b_j| < \infty$. In addition, we assume that $k_0 = [T\tau_0]$ for some $\tau_0 \in (0, 1)$. When $\gamma_1 = \gamma_2$, but $\alpha_1 \neq \alpha_2$, there is a shift in the mean of the long-run equilibrium. This intercept shift is often visualized as a change in the "gap" between two cointegrated series. When $\gamma_1 \neq \gamma_2$, there is a shift in the cointegrating relationship. Let $\beta_1 = (\alpha_1, \gamma_1)'$ and $\beta_2 = (\alpha_2, \gamma_2)'$ and $\delta = (\alpha_1 - \alpha_2, \gamma_1 - \gamma_2)'$. Let $X_t = (1, x_t)'$.

It is evident that the larger the magnitude of a shift, the easier it is to identify the break. The rate of convergence for the estimated break point depends not only on the magnitude of the shift in the coefficients, but also on the magnitude of the regressors. For stationary X_t, the rate depends on the "effective magnitude of shift", $\delta' Q \delta / \sigma_\epsilon^2$, where $Q = E(X_t X_t')$. In this case, it can be shown that $\hat{k} = k_0 + O_p(1)$. This is the best rate that can be achieved for stationary regressors; see, e.g., Bai (1994, 1995). Furthermore, $\hat{k}$ itself is not consistent for k_0, although in terms of the sample fraction, $\hat{k}/T$ converges at a rate of T to τ_0. With I(1) regressors, we shall show that even $\hat{k}$ becomes consistent for k_0. This follows because the "effective magnitude of shift" converges to infinity. More specifically, if $\gamma_1 \neq \gamma_2$, then $a(t) = \delta' E(X_t X_t')\delta / \sigma_\epsilon^2 \to \infty$, as $t \to \infty$. In particular, $a(k_0) \to \infty$ as $T \to \infty$. Consequently, one can estimate the break point more precisely than with I(0) regressors. In particular, $\hat{k} = k_0 + O_p(1)$; see Bai, Lumsdaine and Stock (1994) for a proof. Based on this fact, we can establish a more interesting result. Namely, $P(\hat{k} = k_0) \to 1$.

Theorem 4 *Let $\hat{k}$ denote the least squares estimator of k_0. Assuming that $k_0 = [T\tau_0]$. For the shifted-cointegrating relationship of (16) with $\gamma_1 \neq \gamma_2$, we have*

$$P(\hat{k} = k_0) \to 1.$$

Proof of Theorem 4. Because $\hat{k} = k_0 + O_p(1)$, for any $\epsilon > 0$, there exist an $M < \infty$ such that $P(|\hat{k} - k_0| > M) < \epsilon$. Thus

$$\begin{aligned} P(\hat{k} \neq k_0) &= P(|\hat{k} - k_0| > M) + P(|\hat{k} - k_0| \leq M, \hat{k} \neq k_0) \\ &\leq \epsilon + P(\hat{k} \in D_M) \end{aligned} \tag{17}$$

where $D_M = \{k : |k - k_0| \leq M, k \neq k_0\}$. By definition, $S_T(\hat{k}) \leq \sum_{t=1}^T \varepsilon_t^2$. It follows that the event $\{\hat{k} \in D_M\}$ implies that $\{\min_{k \in D_M} S_T(k) \leq \sum_{t=1}^T \varepsilon_t^2\}$. Thus

$$P\Big(\hat{k} \in D_M\Big) \leq P\Big(\min_{k \in D_M} S_T(k) \leq \sum_{t=1}^T \varepsilon_t^2\Big). \tag{18}$$

We show that the right hand side of (18) converges to zero for every given M. Since D_M is a finite set, it is sufficient to show that for each $k \in D_M$, $P(S_T(k) \leq \sum_{t=1}^T \varepsilon_t^2) \to 0$.

We prove this for $k < k_0$. The case of $k > k_0$ is similar. Define

$$\begin{array}{llllll} Y_{1,k} & = & (y_1, ..., y_k)', & Y_{k,T} & = & (y_{k+1}, ..., y_T)' \\ X_{1,k} & = & (X_1, ..., X_k)', & X_{k,T} & = & (X_{k+1}, ..., X_T)' \\ \varepsilon_{1,k} & = & (\varepsilon_1, ..., \varepsilon_k)', & \varepsilon_{k,T} & = & (\varepsilon_{k+1}, ..., \varepsilon_T)' \\ X_{k,k_0} & = & (X_{k+1}, ..., X_{k_0})', & X^*_{k,T} & = & (X_{k+1}, ..., X_{k_0}, 0, ..., 0)'. \end{array}$$

Furthermore, let $M_1 = I - X'_{1,k}(X'_{1,k}X_{1,k})^{-1}X_{1,k}$ and $M_2 = I - X'_{k,T}(X'_{k,T}X_{k,T})^{-1}X_{k,T}$. Then for $k < k_0$, we have

$$Y_{1,k} = X_{1,k}\beta_1 + \varepsilon_{1,k}$$

$$Y_{k,T} = X_{k,T}\beta_2 + \varepsilon_{k,T} + X^*_{k,T}\delta.$$

The sum of squared residuals $S_T(k)$ is given by

$$\begin{aligned} S_T(k) &= Y'_{1,k}M_1Y_{1,k} + Y'_{k,T}M_2Y_{k,T} \\ &= \varepsilon'_{1,k}M_1\varepsilon_{1,k} + \varepsilon'_{k,T}M_2\varepsilon_{k,T} + 2\delta' {X^*_{k,T}}' M_2\varepsilon_{k,T} + \delta' {X^*_{k,T}}' M_2 X^*_{k,T}\delta \\ &= \sum_{t=1}^{T}\varepsilon_t^2 - \varepsilon'_{1,k}X_{1,k}(X'_{1,k}X_{1,k})^{-1}X'_{1,k}\varepsilon_{1,k} - \varepsilon'_{k,T}X_{k,T}(X'_{k,T}X_{k,T})^{-1}X'_{k,T}\varepsilon_{k,T} \\ &\quad +2\delta' \sum_{t=k+1}^{k_0} X_t\varepsilon_t - 2\delta X'_{k,k_0}X_{k,k_0}(X'_{k,T}X_{k,T})^{-1}X_{k,T}\varepsilon_{k,T} \\ &\quad +\delta'\Big(\sum_{t=k+1}^{k_0} X_tX'_t\Big)\delta - \delta' X'_{k,k_0}X_{k,k_0}(X'_{k,T}X_{k,T})^{-1}X'_{k,k_0}X_{k,k_0}\delta \\ &= \sum_{t=1}^{T}\varepsilon_t^2 + a_T + b_T + c_T + d_T + e_T + f_T \end{aligned}$$

We have used the fact that ${X^*_{k,T}}'\varepsilon_{k,T} = \sum_{t=k+1}^{k_0} X_t\varepsilon_t$ and ${X^*_{k,T}}'X_{k,T} = X'_{k,k_0}X_{k,k_0} = \sum_{t=k+1}^{k_0} X_tX'_t$. For each $k \in D_M$, it is easy to see that a_T, b_T, d_T, and f_T are all $O_p(1)$. Thus

$$S_T(k) - \sum_{t=1}^{T}\varepsilon_t^2 = 2\delta' \sum_{t=k+1}^{k_0} X_t\varepsilon_t + \delta'\Big(\sum_{t=k+1}^{k_0} X_tX'_t\Big)\delta + O_p(1) \tag{19}$$

For each $k < k_0$, the first term on the right hand side of (19) is bounded by $O_p(\sqrt{T})$, whereas the second term is $O_p(T)$ and dominates in magnitude the first term. For example, for $k = k_0 - 1$, the sum involves only one summand, and

$$S_T(k) - \sum_{t=1}^{T}\varepsilon_t^2$$

$$
\begin{aligned}
&= 2\delta' X_{k_0}\varepsilon_{k_0} + \delta'(X_{k_0}' X_{k_0})\delta + O_p(1) \\
&= 2\delta_1\varepsilon_{k_0} + 2\delta_2\Big(\sum_{i=1}^{k_0} e_i\Big)\varepsilon_{k_0} + \delta_1^2 + 2\delta_1\delta_2\Big(\sum_{i=1}^{k_0} e_i\Big) + \delta_2^2\Big(\sum_{i=1}^{k_0} e_i\Big)^2 + O_p(1) \qquad (20) \\
&= O_p\Big(\delta_2 \sum_{i=1}^{k_0} e_i\Big) + \Big(\delta_2 \sum_{i=1}^{k_0} e_i\Big)^2.
\end{aligned}
$$

The last term can be rewritten as $T\delta_2^2(\frac{1}{\sqrt{T}}\sum_{i=1}^{[T\tau_0]} e_i)^2$, which dominates the first term and converges to positive infinity with probability approaching 1. This implies that for any $\epsilon > 0$,

$$
P\Big(S_T(k) - \sum_{t=1}^{T} \varepsilon_t^2 \leq 0\Big) < \epsilon
$$

for large T. Combining with (17) and (18), we have, for every $\epsilon > 0$, $P(\hat{k} \neq k_0) < 2\epsilon$ for all large T. $\square$.

When only the intercept has a break ($\alpha_1 \neq \alpha_2$, $\gamma_1 = \gamma_2$), $\hat{k}$ is no longer consistent for k_0, even though $\hat{k}/T$ still converges to τ_0 at a rate of T. This is because the "effective magnitude of shift" stays bounded. The lack of consistency can also be seen from (20). When $\delta_2 = 0$, $S_T(k) - \sum_{t=1}^{T}\varepsilon_t^2 = 2\delta_1\varepsilon_{k_0} + \delta_1^2 + O_p(1)$, which cannot be guaranteed to be positive.

The underlying reason for $\hat{k}$ being consistent for k_0 is not the I(1) regressor per se. Roughly speaking, if $\hat{k} = k_0 + O_p(1)$ holds, then for any set of regressors X_t such that the second term (19) converges to infinity and dominates the first term, $\hat{k}$ will be consistent for k_0. In particular, this is true for polynomials regressions. For simplicity, we consider the following broken-trend model:

$$
y_t = \begin{cases} \alpha_1 + \gamma_1 t + \varepsilon_t & t = 1, 2, ..., k_0 \\ \alpha_2 + \gamma_2 t + \varepsilon_t & t = k_0 + 1, ..., T \end{cases} \qquad (21)
$$

where ε_t is a linear process such that $\varepsilon_t = \sum_{j=0}^{\infty} a_j \eta_{t-j}$ with η_t a sequence of martingale differences and $\sup_i E\eta_i^4 < \infty$, and $\sum_{j=0}^{\infty} j|a_j| < \infty$.

Theorem 5 *For the broken trend model (21), assume $\gamma_1 \neq \gamma_2$. Let $\hat{k}$ be the least squares estimator of k_0 with $k_0 = [T\tau_0]$. Then*

$$
P(\hat{k} = k_0) \rightarrow 1, \qquad \textit{as } T \rightarrow \infty
$$

Proof of Theorem 5. Let $X_t = (1, t)'$. Then the proof of Theorem 4 up to equation (19) can be copied here. The right hand side (19) still converges to positive infinity (with probability 1) for the newly defined X_t for each $k \in D_M$. For example, for $k = k_0 - 1$,

$$\begin{aligned} S_T(k) - \sum_{t=1}^{T} \varepsilon_t^2 &= 2\delta' X_{k_0} \varepsilon_{k_0} + \delta'(X'_{k_0} X_{k_0})\delta + O_p(1) \\ &= 2\delta_1 \varepsilon_{k_0} + 2\delta_2 k_0 \varepsilon_{k_0} + \delta_1^2 + 2\delta_1\delta_2 k_0 + (\delta_2 k_0)^2 + O_p(1) \qquad (22) \\ &= O_p(\delta_2 k_0) + (\delta_2 k_0)^2 \\ &= O_p(\delta_2 T \tau_0) + (\delta_2 T)^2 \tau_0^2 \end{aligned}$$

It follows that the second term above dominates the first and converges to positive infinity. This implies that, for each $k \in D_M$, $P(S_T(k) - \sum_{t=1}^{T} \varepsilon_t^2 \leq 0) \to 0$. □

Although Theorem 5 is not explicitly presented in the literature, it is not unexpected. In Bai (1994, 1995), the linear trend is written as $\frac{t}{T}$ and it is proved that $\hat{k} = k_0 + O_p(1)$ (but $\hat{k}$ is not consistent for k_0). If we rewrite the broken trend model (21) with the linear trend expressed in the format $\frac{t}{T}$, then the new slope coefficients become $T\gamma_1$ for $t \leq k_0$ and $T\gamma_2$ for $t > k_0$. The magnitude of shift will be $T(\gamma_2 - \gamma_1)$. Thus, if the model is cast in the framework of Bai (1994, 1995), one is essentially assuming an unbounded magnitude of shift. In this sense, the consistency of $\hat{k}$ for k_0 is not surprising.

The results of Theorem 4 and Theorem 5 still hold even if we allow the magnitude of shift to converge to zero. Let $\delta_{2,T} = \gamma_2 - \gamma_1$.

Theorem 6 *For the cointegrated regression model (16), assume $\sqrt{T}\delta_{2,T} \to \infty$. For the broken trend model (21), assume $T\delta_{2,T} \to \infty$. Under these assumptions, even if $\delta_{2,T} \to 0$, the estimated break point $\hat{k}$ is still consistent for k_0. That is, $P(\hat{k} = k_0) \to 1$.*

Proof of Theorem 6. Under the assumed magnitude of shift, we still have $\hat{k} = k_0 + O_p(1)$; see Bai (1994, 1995) and Bai, Lumsdaine and Stock (1994).[2] Consider the

[2] In fact, the assumed magnitude of shift is stronger than necessary. For example, for the broken-trend model, $\delta_{2,T}/\sqrt{T} \geq c > 0$ is sufficient for $\hat{k} = k_0 + O_p(1)$. But the assumption is necessary for $\hat{k}$ to be consistent for k_0.

cointegrating regression (16). All that is needed is to prove (19) converges to positive infinity with probability 1. As before, we consider $k = k_0 - 1$. By (20),

$$S_T(k) - \sum_{t=1}^{T} \varepsilon_t^2 = O_p\Big(\sqrt{T}\delta_{2,T}\frac{1}{\sqrt{T}}\sum_{i=1}^{k_0} e_i\Big) + \Big(\sqrt{T}\delta_{2,T}\frac{1}{\sqrt{T}}\sum_{i=1}^{k_0} e_i\Big)^2 \tag{23}$$

Because $\frac{1}{\sqrt{T}}\sum_{i=1}^{k_0} e_i$ converges to a normal random variable and $\sqrt{T}\delta_{2,T} \to \infty$, the right hand side of (23) converges to positive infinity. The proof for the broken trend model is similar and thus omitted. □

References

[1] Bai, J. (1994) Estimation of Structural Change based on Wald Type Statistics. Working Paper No. 94-6, Department of Economics, M.I.T.

[2] Bai, J. (1995) Least Absolute Deviation Estimation of a Shift. *Econometric Theory* 11, 403-436.

[3] Bai, J., R.L. Lumsdaine, and J.H. Stock (1994) Testing for and Dating Breaks in Integrated and Cointegrated Time Series. Manuscript, Kennedy School of Government, Harvard University.

[4] Gregory, A.W. and B.E. Hansen (1996) Residual-based tests for cointegration in models with regime shifts. *Journal of Econometrics* 70 99-126.

[5] Hansen, B.E. (1992) Tests for parameter instability in regressions with I(1) processes. *Journal of Business and Economic Statistics* 10, 321-335.

[6] Nunes, L.C., C.M. Kuan, P. Newbold (1995) Spurious Break. *Econometric Theory,* 11, 736-749.

[7] Quintos, C.E. and P.B.C. Phillips (1993) Parameter constancy in cointegrating regressions. *Empirical Economics* 18, 675-703.